DOCUMENT RESUME

ED 185 120 TM 800 161

AUTHOR Kokosh, John
TITLE Finding and Developing Moderators and Directional Keys by Regression Analysis.
PUB DATE Sep 79
NOTE 16p.: Paper presented at the Annual Meeting of the American Psychological Association (87th, New York, NY, September 1-5, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance; *Discriminant Analysis; Item Analysis; Multiple Regression Analysis; *Predictor Variables; *Statistics
IDENTIFIERS *Moderator Variables

ABSTRACT

A procedure for rapid screening of variables as potential moderators is presented and discussed. A moderator is defined as any variable which can be used to identify differentially predictable persons; or defined statistically by stating that if a predictor and a moderator are each divided into three or more categories and used as independent variables to predict a criterion, then an F test on interaction effects will determine the presence of a moderator variable. However, in sets of data based on large numbers of variables, it is difficult to discover moderators. The method proposed begins by arranging the subjects in four quadrants, according to their scores on a predictor and the criterion, and coding the hit and miss cells as dummy variables (zero and one, respectively) and then computing the point-biserial correlation of the remaining variables with this dummy variable. In addition, one of the cells coded one may be changed to minus one, to represent underprediction, in order to develop a directional key for the moderators. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185120

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John Kokosh

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Finding and Developing Moderators and Directional Keys by Regression Analysis

John Kokosh

DePaul University

Paper presented at the annual meeting of the American Psychological Association, New York, New York, September 1979.

IM800 161

Abstract

A modification of the Ghiselli procedure, based on discriminant analysis, is presented for rapid screening of variables as potential moderators which can then be used in multiple regression models to improve the prediction of human performance. In addition, if a pool of items is available for such prediction research, then the same research design can also be used to develop both a moderator variable and a "directional key", all of which can be tested and used in one multiple regression equation. This methodology is presented to obviate some weaknesses inherent in the Ghiselli procedure and, at the same time, maximize the usefulness of one's prediction data.

Finding and Developing Moderators and Directional Keys by Regression Analysis

One of the major criticisms of research on moderator variables is that there are no systematic procedures for finding moderators without sub-grouping samples and then computing many correlation coefficients for each group (Banas, 1964; and Zedeck, 1971b). The purpose of this paper is to show how discriminant analysis can be used for finding moderators which can then be used in a multiple regression equation to improve the prediction of human performance. In addition, if a pool of items is available for such prediction research, then discriminant analysis can also be used to "develop" a moderator variable by item analysis. Finally, the same research design will make it possible to develop a "directional key" which will be orthogonal or uncorrelated with the predictor. These three kinds of variables, the predictor, the moderator, the directional key, and the cross-product of the predictor and the moderator can be tested and used in one multiple regression equation to predict a criterion.

Brief Literature Review

Definitions

Some of the more mundane problems that have emerged from research on moderator variables are due simply to a misunderstanding as to what is and what is not a moderator variable. The simplest definition of a moderator is that it is any variable which can be used to identify differentially predictable persons. This can be a categorical variable, such as sex (Seashore, 1962; Schmitt, Mellon, & Bylenga, 1978), or a continuous variable, such as anxiety (Malnig, 1959). Either type of moderator variable can be used in a multiple regression equation as shown by Cohn (1968; 1978);

that is, a categorical moderator can be coded as a dummy variable and used in the same way as a continuous variable to predict a criterion. A good statistical definition of a moderator variable is in terms of interaction effects in an analysis of variance (ANOVA) model. If a predictor $\underline{X}_1$ and a moderator $\underline{X}_2$ are each broken down into three or more categories and used as independent variables, and the criterion $\underline{Y}$ is used as the dependent variable, then an $\underline{F}$ test on interaction effects will determine the presence of a moderator variable.

Saunders' (1956) moderated regression equation, shown as Equation 1, can also be used to define and evaluate a moderator variable. However, the cross-product term, $\underline{B}_3\underline{X}_1\underline{X}_2$, will only account for linear - times - linear interaction effects. The ANOVA model, on the other hand, provides

$$\underline{Y}' = \underline{A} + \underline{B}_1\underline{X}_1 + \underline{B}_2\underline{X}_2 + \underline{B}_3\underline{X}_1\underline{X}_2 \qquad (1)$$

an omnibus $\underline{F}$ test on interaction effects, which can be further broken down into linear and non-linear interaction components. The researcher can then write an appropriate regression equation which will take into account the more complex kinds of interaction effects.

Moderator variables can also be developed by item analysis. One procedure, used by Ghiselli (1956), involves converting the predictor ($\underline{X}$) and criterion ($\underline{Y}$) into standard ($\underline{z}$) scores, subtracting one from the other to obtain the deviation scores, or deviations from what Ghiselli calls a "line of relations", and then performing an item analysis against the absolute deviations ("$\underline{d}$") from that line. Thus, subjects who obtain low scores on this new scale are closer to the line of relations and are, therefore, more predictable than those who obtain high scores.

Mention should be made of two kinds of variables which are not moderators. Abrahams and Alf (1972a), in an excellent critique of moderator variables

research, point out that the keys developed by Hobert (1965), and reported in Hobert and Dunnette (1967), do not qualify as moderator variables, for reasons too complicated to explain here; one rather simple reason is that such keys reflect main effects in ANOVA terms, rather than interaction effects, and can thus be tested and used in a conventional multiple regression equation.

Another kind of a variable, which is not a moderator, is the "deviate key" developed by Neidt and Malloy (1954) which will identify the overpredicted and the underpredicted subjects. This, too, can be used in a multiple regression equation as was correctly done by Neidt and Malloy. However, some researchers, including Jones (1969), Zedeck (1971a), and Raubenheimer and Tiffin (1971), have used the deviate technique to either develop or find "moderator variables" without realizing that such variables were predictors which could be used in a conventional multiple regression equation (no cross-product terms) and without subgrouping samples to improve prediction. Abrahams and Alf (1972a) demonstrate algebraically that such keys or variables are predictors and not moderators.

The Search for Moderators

One of the more obvious problems or questions regarding moderator variables is how to find them without subgrouping samples, on the basis of each potential moderator, and then calculating many separate validity coefficients. French (Note 1) attempted to develop a program to identify moderators, from among many variables, which would eliminate the need to compute many correlations. Using 42 variables as potential moderators, and pairing each with the criterion in a 3 x 3 contingency table, with each variable trichotomized, he systematically searched for effective moderators. But the program was not successful, according to Zedeck (1971b), apparently

because of distortions with the 3 x 3 indexing technique.

The Ghiselli (1956) procedure for developing moderators, described earlier, can also be used for finding or detecting moderators from among many variables. Both Banas (1964) and Zedeck (1971a) correlated potential moderators with "d"scores. Zedeck had no success with the Ghiselli procedure in finding moderator variables. Banas appears to have had some success with this procedure, but this writer was not able to determine how successful Banas had been.

Abrahams and Alf (1972a) point out some potential problems with the Ghiselli procedure. There is a tendency for subjects who are either very high or very low on the predictor variable to obtain high "d" scores, thus identifying them erroneously as "unpredictable", which is rather unfortunate since the very high predictor and the very low predictor subjects are the ones who are the most and least desired, respectively, in most selection programs. The reason why the people at the two extreme ends of the predictor tend to have high "d" scores, according to Abrahams and Alf, is because they are less reliably measured. Consequently, the Ghiselli procedure may not be an effective way for finding moderator variables even though it may be an adequate procedure for developing them by item analysis.

Another potential problem with the Ghiselli procedure is that when there is a restriction of range on either the predictor, or the criterion, or both, then the low scoring subjects on either variable will tend to be identified as "unpredictable". The problem here is similar to what occurs at either end of the bivariate distribution, described by Abrahams and Alf, but the restriction of range situation will probably have a more serious adverse effect on the Ghiselli procedure than the attenuation problem. Thus, it appears that efficient procedures are still needed for finding and/or developing moderator variables.

Finding Moderators

The following procedure for finding moderator variables, based on discriminant analysis, is a modification of the Ghiselli procedure and will reduce the tendency of erroneously identifying the very high scoring and the very low scoring subjects as "unpredictable". The method is based on the quadrant format similar to that used by Hobert (1965), shown in Figure 1, but the method of analysis is quite different from that of Hobert.

Insert Figure 1 about here

The discriminant analysis procedure is as follows: (1) arrange the subjects into four cells, or quadrants; (2) code the subjects in the two "hit" cells as zero, and those in the two "miss" cells as one; and (3) compute a point-biserial correlation between each potential moderator and the dichotomous (0, 1) criterion. This analysis can be easily done either with a discriminant computer program or with a conventional regression program by coding the "hit" and "miss" cells as dummy variables (0, 1) and computing point-biserial correlations against them (McNemar, 1969, pp. 234-235; Kerlinger & Pedhazur, 1973, pp. 237-238).

Several variations on the model just described can be followed when the correlation between the predictor and the criterion is high resulting in a rather uneven distribution of subjects between the "hit" and "miss" cells. For example, when the validity coefficient is about .40 or higher, the predictor and the criterion can each be trichotomized, the subjects in the three "hit" cells along the right diagonal can be coded as zero, and the subjects in the six remaining or "miss" cells coded as one; this should yield a more even number of subjects in each group or 0, 1 category.

Another variation on this correlational procedure would be to break the sample down into, say, twenty-five cells, with five categories on the predictor and five on the criterion. In addition, the cells could be coded on a continuous five-point scale, assigning a code of zero to the five "hit" cells along the right diagonal, and a code of one to four for the remaining cells, depending on their distance from the diagonal. But at this point the design of the analysis begins to resemble the Ghiselli procedure and, thus, will tend to have the same drawbacks including the tendency to identify the very high and the very low predictor subjects as "unpredictable".

Mention should be made of some useful features of the discriminant analysis search method which applies to the Ghiselli procedure as well. First, one can identify a curvilinear correlation between a predictor and the criterion, but only a quadratic function, simply by correlating the continuous predictor with the 0, 1 criterion. Second, the data analysis can be done in a stepwise computer program so that "joint moderators" (Frederiksen & Melville, 1954; Frederiksen & Gilbert, 1960; Striker, 1966; and Zedeck, Cranny, Vale, & Smith, 1971) can be identified. Finally, the results of the analysis, in the form of correlation coefficients, will indicate the direction of predictability for any given moderator. Thus, a positive correlation between a moderator and the 0, 1 criterion would mean that the higher a score a person has on a moderator, the less predictable that person appears to be with a given predictor.

Developing Moderators and Directional Keys

The quadrant format can also be used to develop a moderator variable and a directional key, similar to the "deviate key" developed by Neidt and Malloy (1954), if a pool of items is available to do the job. First, the moderator key can be developed by correlating each item with the dichotomous 0, 1 criterion, where the two hit cells are coded zero, and the two miss cells as

as one, as shown in Figure 1. Second, the items for the directional key can be found by computing correlations between each item and the three groups of the quadrant format, with the underpredicted group (low predictor, high criterion) coded one, the two hit cells coded as zero, and the overpredicted (high predictor, low criterion) coded as minus one. Items for the two keys would be selected if they show a significant correlation with their respective criteria, but no item should be used in both scales.

The procedure just described would enable the researcher to operationalize an idea proposed by Dunnette (1972), namely to develop a directional key as well as a moderator variable by item analysis for a given prediction problem. In addition, this procedure is simpler and less troublesome from a statistical standpoint than the methodology proposed by Dunnette which was criticized by Abrahams and Alf (1972b). Finally, if a moderator variable and a directional key are developed by item analysis, both keys can be used in one multiple regression equation. Such a prediction model would include the original predictor, the directional key, the moderator key, and the cross-product of the predictor and the moderator. No subgrouping would be needed or desired, to develop additional predictors, since the pool of available items would have already been thoroughly mined for its predictive value. The whole system would, of course, have to be either replicated or cross-validated. Better still would be to use three samples: two samples to develop the moderator and the directional key (i.e., by replication), and a third sample to cross-validate the complete model.

The results of such research should help to clarify some basic questions about the nature and usefulness of moderator variables. It should also make some contribution to psychometric theory and the art and science of test development. Finally, the development of moderator variables and directional

keys may prove to be one effective way that psychologists can go beyond the prediction ceiling of about .50, noted by Hull (1928) and Ghiselli (1955). Such a prediction strategy would, at least, make the most efficient use of one's prediction data.

Reference Notes

1. French, J. W. A machine search for moderator variables in massive data. Princeton, New Jersey: Educational Testing Service, ONR Tech. Rep., 1961.

References

Abrahams, N. M., and Alf, E., Jr. Pratfalls in moderator research. Journal of Applied Psychology, 1972, 56, 245-251. (a)

Abrahams, N. M., and Alf, E., Jr. Reply to Dunnette's "Comments on Abrahams Alf's 'Pratfalls in moderator research'". Journal of Applied Psychology, 1972, 56, 257-261. (b)

Banas, P. A. An investigation of transituational moderators. Unpublished doctoral disseration, University of Minnesota, 1964.

Cohen, J. Multiple regression as a general data analytic system. Psychological Bulletin, 1968, 70, 426-443.

Cohen, J. Partialed products are interactions; partialed powers are curve components. Psychological Bulletin, 1978, 85, 858-866.

Dunnette, M. D. Comments on Abrahams and Alf's "Pratfalls in moderator research". Journal of Applied Psychology, 1972, 56, 252-256.

Frederiksen, N. O., and Gilbert, A. C. F. Replication of a study of differential predictability. Educational and Psychological Measurement, 1960, 20, 759-767.

Frederiksen, N. O., and Melville, D. S. Differential predictability in the use of test scores. Educational and Psychological Measurement, 1954, 14, 647-656.

Ghiselli, E. E. The measurement of occupational aptitude. Berkeley: University of California Press, 1955.

Ghiselli, E. E. Differentiation of individuals in terms of their predictability. Journal of Applied Psychology, 1956, 40, 374-377.

Hobert, R. D. Moderating effects in the prediction of managerial success from psychological test scores and biographical factors. Unpublished doctoral dissertation, University of Minnesota, 1965.

Hobert, R. D., and Dunnette, M. D. Development of moderator variables to enhance the prediction of managerial effectiveness. Journal of Applied Psychology, 1967, 51, 50-64.

Hull, C. L. Aptitude testing. Yonkers, New York: World Books, 1928.

Jones, I. A. The CPI as a moderator variable in the prediction of grade point average from ACT scores. (Doctoral dissertation, Baylor University) Ann Arbor, Michigan: Univer. Microfilms, 1969, (Order NO. 69-17, 841).

Kerlinger, F. N., and Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart and Winston, Inc., 1973.

Malnig, L. R. Differential predictability of academic achievement from aptitude tests administered under psychological stress to subjects with varying amounts of anxiety. Unpublished doctoral dissertation, New York University, 1959.

McNemar, Q. Psychological statistics. (4th ed.) New York: Wiley, 1969.

Neidt, C. O., and Malloy, J. P. A technique for keying items of an inventory to be added to an existing test battery. Journal of Applied Psychology, 1954, 38, 308-312.

Raubenheimer, I. van W., and Tiffin, J. Personnel selection and the prediction of error. Journal of Applied Psychology, 1971, 55, 222-233.

Saunders, D. R. Moderator variables in prediction. Educational and Psychological Measurement, 1956, 16, 209-222.

Schmitt, N., Mellon, P. M., and Bylenga, C. Sex differences in validity for academic and employment criteria, and different types of predictors. Journal of Applied Psychology, 1978, 63, 145-150.

Seashore, H. C. Women are more predictable than men. Journal of Counseling Psychology, 1962, 9, 261-270.

Stricker, L. J. Compulsivity as a moderator variable: A replication and

extension. Journal of Applied Psychology, 1966, 50, 331-335.

Zedeck, S. Identification of moderator variables by discriminant analysis in a multipredictable group validation model. Journal of Applied Psychology, 1971, 55, 364-371. (a)

Zedeck, S. Problems with the use of moderator variables. Psychological Bulletin, 1971, 76, 295-310. (b)

Zedeck, S., Cranny, C. J., Vale, C. A., and Smith, P. C. Comparison of "joint moderators" in three prediction techniques. Journal of Applied Psychology, 1971, 55, 234-240.

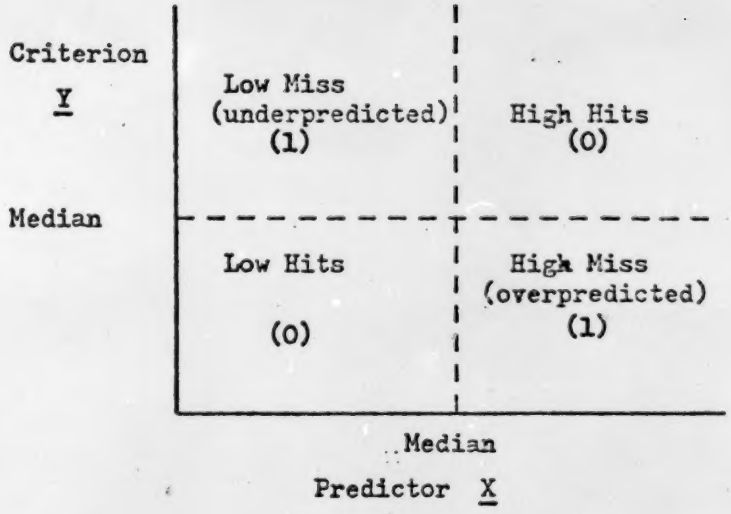

Figure 1. Diagram of the quadrant format with the predictor X and the criterion Y each dichotomized at their respective median scores.